IdeaBot: Investigating Social Facilitation in Human-Machine Team Creativity

ANONYMOUS AUTHOR(S)

The present study investigates how human subjects collaborate with a computer-mediated chatbot in creative idea generation tasks. In three text-based between-group studies, we tested whether the perceived identity (i.e.,whether the bot is perceived as a machine or as a human) or the conversational style of a teammate would moderate the outcomes of participants' creative production. In Study 1, participants worked with either a chatbot or a human confederate. In Study 2, all participants worked with a human teammate but were informed that their partner was either a human or a chatbot. Conversely, all participants worked with a chatbot in Study 3, but were told the identity of their partner was either a chatbot or a human. We investigated differences in idea generation outcomes and found that participants consistently contributed more ideas and with ideas of higher quality when they perceived their teamworking partner as a bot. Furthermore, when the conversational style of the partner was robotic, participants with high anxiety in group communication reported greater creative self-efficacy in task performance. Finally, whether the perceived dominance of a partner and the pressure to come up with ideas during the task mediated positive outcomes of idea generation also depends on whether the conversational style of the bot partner was robot- or human-like. Based on our findings, we discussed implications for future design of artificial agents as active team players in collaboration tasks.

CCS Concepts: • **Human computer interaction (HCI)** → HCI theory, concepts and model.

Additional Key Words and Phrases: chatbot, teamworking, social facilitation, creativity, idea generation

ACM Reference Format:
Anonymous Author(s). 2020. IdeaBot: Investigating Social Facilitation in Human-Machine Team Creativity. *ACM Trans. Graph.* 37, 4, Article 111 (August 2020), 23 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Creativity and collaboration have played a critical role in human history. However, as a research area, creativity is a enigmatic topic, and some of its most fundamental inquiries, such as how we can enhance our creativity, remain open and unsolved. Studying creativity in teams only adds complexity to the topic, but this is the typical setting in which we conduct brainstorming and generate extraordinary ideas that we can hardly come up with solo. Existing literature in group communication and organizational behaviors describes positive and negative effects of teamwork on individuals' productivity and creativity [3, 19, 21, 38]. For instance, while our teammates may bring diverse perspectives to the table, social pressures may result in groupthink and thus hinder productive outcomes of teamworking [38]. Partners can distract one from focusing on the task at hand or hinder one from proposing different ideas against an opinion leader [37]. We may have little control over our human collaborators, but the possibility for designing computer-mediated team players is infinite. This may allow us to attain some of the benefits of teamworking while removing some negative effects of social pressures. With the burgeoning use of artificial intelligence (AI) in communication applications, recent research

Author's address: Anonymous Author(s).

Manuscript submitted to ACM

has shown examples of AI assistants facilitating trustful and efficient communication between human teammates [24, 25, 27]. Taking a step further, we ask how an autonomous agent may contribute as an active team player, instead of a passive assistant, and specifically, how such direct collaboration with an artificial partner may influence creative production. Thus, in this study, we investigate how people may work with a chatbot teammate on a idea generation task. The goal of the present research is to better understand the effect of the *perceived identity* and the *conversational styles* of collaborators on team idea generation outcomes. Our primary research question is thus:

RQ1: Can creative ideation be improved by working with a collaborator who ***is****, or* ***is perceived to be****, a bot?*

Understanding the effect of perceived *identity* can help solidify the future role of artificial agents in creative team collaboration. For instance, if participants show reluctance to collaborate with an non-human teammates, future research and application can focus on adopting autonomous agents as facilitators and reduce direct interaction between humans and machine teammates. in addition, regardless of the role of future AI agents in teamworking, understanding the effect of *conversational style* informs design implications for future bots. That is, are there more positive effects when a bot partner acts in an robotic or human-like fashion? While previous literature has suggested bots with more anthropomorphic cues can elicit more interpersonal interaction from its human users [8, 9], we are interested in whether a machine-sounding agent can also be beneficial in certain cases.

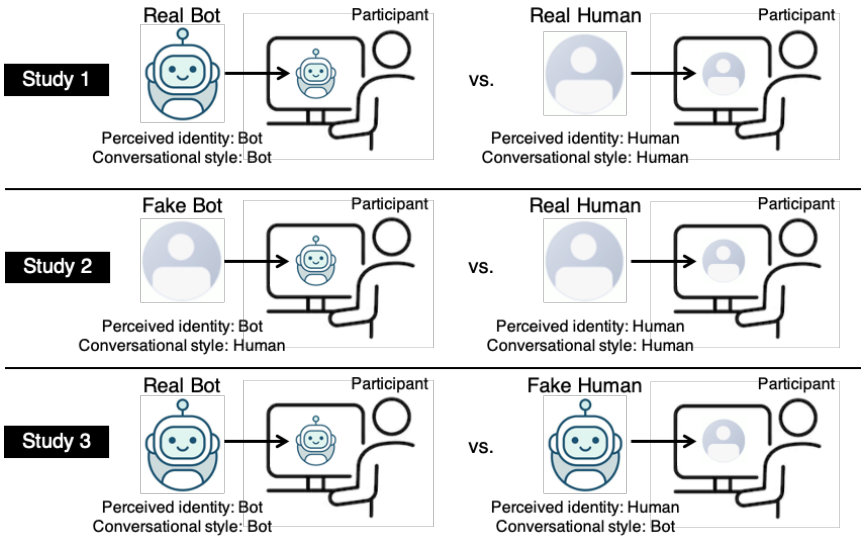


Fig. 1. Experiment design of the three studies in the present research.

In order to tackle these inquiries, we review research across the fields of social psychology, communication, and human-robot interaction (HRI). We design and conduct three text-based studies where participants were paired with either a human or a bot collaborator to generate creative ideas for a brainstorming task. In Study 1, participants either interacted with a human (*real human*) or a bot partner (*real bot*) to work on a brainstorming task. In Study 2, all participants worked with a human confederate but were told that their partner was either a human (*real human*) or a bot (*fake bot*). In Study 3, all participants brainstormed with a chatbot but were told that their partner was either a human (*fake human*) or a bot (*real bot*). We perform systematic comparisons to understand whether individuals demonstrate different cognitive and behavioral patterns when collaborating with human versus artificial confederates. Specifically,

we evaluate the effect of perceived identity and conversational styles respectively on four aspects: (1) task success and idea generation outcomes, (2) creative self-efficacy, (3) perception of teamworking partners, and (4) conversational experiences (e.g., pauses, attentiveness, expressiveness, and perceived pressure in a conversation). Our results not only show promising potential for autonomous agents to serve as active team players but also offer insights into what types of users may benefit from working with a bot partner and which conversing styles of the bot may make the greatest impact. Our contributions are as follows:

(1) Emphasizing the role of artificial agents as active team players (instead of passive facilitators), we examine the effect of collaborating with a bot vs. with a human partner on creative ideation.
(2) Using the theoretical framework of Social Faciliation and Distraction-Conflict Theory, we test the effects of perceived identity and conversational style on human-machine teamworking and creativity.
(3) We examine how individual differences may moderate the outcomes of teamworking and idea generation with an autonomous agent.
(4) We discuss the design implications for computer-mediated agents performing unstructured tasks (i.e., brainstorming) and propose avenues for future research in human-machine teamworking and creativity.

2 BACKGROUND AND RELATED WORK

2.1 Social Facilitation and Inhibition: The Making and Breaking of Teams

Research on teamwork, especially research examining how humans and machines collaborate, has built heavily on *Social Facilitation Theory* [57]. As one of the oldest theories in social psychology, social facilitation theory suggests the presence of others can motivate and improve individuals' performance. Subsequent research has suggested the effect of social facilitation can be decomposed into three components, including the effect of *mere presence* [1, 56, 59], *cognitive-behavioral cues* [13, 46], and *social cues* [10, 12, 22]. Specifically, the mere presence effect posits that individuals can be influenced by the mere presence of others to perform better. On top of that, the performance of others can drive one's own performance by eliciting a sense of competition or peer pressure. Social cues, on the other hand, relate to evaluative apprehension, when feedback influences one's performance [6, 52].

Despite its rich source of cues and motivations, teamwork does not guarantee positive outcomes. Working in teams can also inhibit creativity [22, 38]. To understand the dynamics of social presence, Sanders and Baron [2, 43] posited the Distraction-Conflict framework, suggesting that individuals' cognitive capacity is divided with the presence of others. Participant attention is divided between performing their task and monitoring the behaviors of their partners. When attention is evenly divided, this conflict serves as a cognitive stimulus, allowing individuals to be more alert and more engaged while participating in the group activity. However, when either of the two sources overrides the other (either when a person is overly concerned about other team members' responses to his/her performance, or when s/he ignores others' contributions and focuses only on completing the task) such conflict can distract a team from producing favorable outcomes [2, 5, 43].

2.1.1 The effect of anxiety in group communication. Whether the social presence of teammates facilitates or inhibits task success can be affected by individual characteristics in team settings, such as being timid, thoughtful, sociable, or judgmental when interacting with teammates. In particular, individual differences exist in the extent to which people feel anxious in group communication [20, 39, 48]. Specifically, former literature [48] classified participants' patterns of interpersonal interaction in teams into five categories: Individuals in Group 1 are bold and demonstrate verbal intrusiveness in group conversation; those in Group 2 are more reserved and tend to keep a distance from

other team members; Group 3 members typically feel more relaxed and show a higher degree of spontaneity, Group 4 members tend to be expressive and critical; and finally, individuals in Group 5 consistently demonstrated nervousness and anxiety in front of their teammates, which inhibited their own performance. This last phenomenon in group communication resonates with the Distraction-Conflict framework. Since the cognitive capacity of highly anxious people is overwhelmed by concerns and worries about others, anxious people are unable to engage fully with the task at hands.

In classic human-computer interaction (HCI) literature, Nass and colleagues posited that humans were also influenced by social feedback when interacting with computer-mediated agents [34]. Subsequent research further found the degree of concern and evaluation apprehensiveness toward artificial agents can be influenced by the number of anthropomorphic features [31, 40]. That is, the more human-like characteristics an artificial confederate exhibited, the more concern participants demonstrated toward their partner. In this regard, we suspect the effect of evaluative apprehension on anxious participants' cognitive distraction will be relatively mild when a person collaborates with a partner whom they believe to be artificial, or, with a partner who demonstrates fewer human-like behavioral cues. However, this hypothesis has not yet been systematically tested in human-robot teamwork. Therefore, we propose our second research question as:

RQ2: How do individual levels of anxiety in group communication moderate teamwork and idea generation with human and bot confederates?

2.1.2 The effect of perception toward team partners and conversational experiences. In a teamwork setting, individuals' performance is also commonly affected by their partners. In particular, research in group communication has found that individuals are susceptible to cues displayed or influenced by their partners in teamwork settings, including *dominance*, *productivity*, and *social evaluation* [12, 35, 38]. A dominating teammate may result in opinion leadership and groupthink (i.e., the entire group passively agrees with opinions of a single contributor). The concept of dominance can be quantitatively measured by the portion of time a person speaks in a conversation or the amount of text delivered in an computer-mediated conversation. Whether the person has served as an opinion leader resulting in groupthink that complies with his/her opinions is also a measure of dominance [33, 35]. Other studies have examined the effect of group members' productivity and found that a less motivated or less productive group member may result in social loafing, lowering the productivity of an entire team [22, 44, 47, 54]. Finally, social psychologists suggested individuals may experience evaluative apprehension when trying to understand groupmates' feedback. Attention to social evaluation may then distract them from their task performance [7, 16, 52]. Therefore, social evaluation has also been found to mediate the effect of social facilitation with interactive partners. To sum up, the current study asks:

RQ3: How do perceptions of team partners and conversational experiences mediate teamwork and idea generation with bot confederates?

2.2 Former Research in Teamwork with Non-human Confederates

Pioneering work in Human-Robot Interaction (HRI) has examined the potential of working with robots to mitigate some of the above-mentioned issues [11, 23, 36, 42]. Unlike the contributions of human collaborators, the pace and content of non-human agent responses can be fine-tuned to make other groupmates more comfortable and engaged [28]. Former research has consistently revealed that, when interacting with artificial collaborators, participants also demonstrated certain behavioral patterns similar to how they would interact with human confederates [11, 23, 42]. Classic HRI literature describes the concept of the "social robot," which suggests that the more humanoid features

that are embedded in a bot (e.g., facial expressions, humanoid appearances), the more likely individuals are to view a machine-mediated agent as a human [8, 9]. Various studies have examined the effect of autonomous agents' physical appearance and concluded that the degree of embodiment can mediate the level of social facilitation [23, 26, 36]. That is, when the confederate was presented in a highly embodied form, such as in a humanoid robot, the social facilitation effect tended to be stronger. Conversely, when a non-human confederate appears in a low-level embodiment format, such as an animated image on the screen, participants were less likely to benefit from social facilitation. Ongoing work in the field found robots in teams can facilitate interpersonal interaction and alleviate conflicts among other human teammates [24, 49, 55]. More recently, scholars have proposed the ideas that the tireless, constantly productive nature of robots may encourage other human teammates in a group to contribute continuously as well [4, 55].

In our literature review, we also found several gaps in the existing literature. First of all, the majority of former studies treated non-human agents as *passive others*, where artificial agents were present in team settings but offered minimal amount of interaction [23, 32, 41, 42, 58, 60]. Therefore, though findings of these experiment provided insights into the mere presence effect of non-human confederates, our understanding toward their cognitive-behavioral and social cues remains limited. On top of that, given the facilitating or even passive role of bots in former research, evaluation of experiment outcomes often focused on task success and/or completion. As a result, there were rather limited research studying human-robot teamworking through unstructured tasks, and the adoption of a idea generation task was even less common. Hence, our understanding toward human-robot teamworking in creative idea remains underexplored.

Also, while a considerable amount of former research has been dedicated to the refinement of robot design, a common approach in empirical design was to compare various versions of bots. By contrast, a core interest of the present study lies in comparing the difference between how individuals interact with human versus non-human collaborators. Abundant literature in HRI has suggested when individuals activate a distinct cognitive system when interacting with robotic entities [14, 61]. Therefore, in the present study, we investigate the teamworking outcomes when participants collaborate with confederates that are either perceived as a bot or as a human, while assessing how the informed and perceived artificiality of the collaborative entity may elicit an effect. In summary, the conceptual framework of the present study is illustrated in 2.

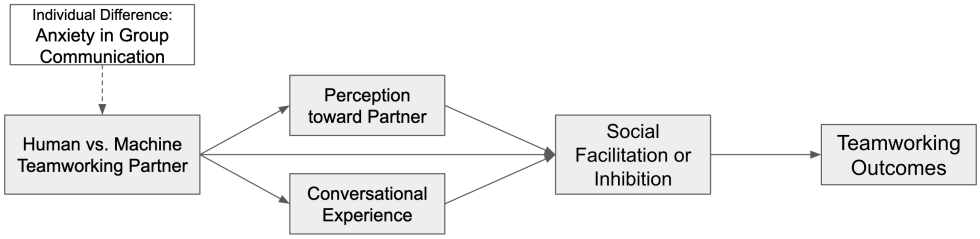


Fig. 2. A conceptual framework of the present study.

Our contributions are thus threefold. First of all, while former human-machine collaboration research mostly adopted artificial confederates as passive audience (i.e., machines did not directly interact with individuals) [23, 32, 41, 42, 58, 60], the current study will investigate non-human agents as interactive co-actors in a team idea generation task (i.e., machines will directly participate in the idea generation process). Secondly, existing human-robot interaction (HRI) and human-computer interaction (HCI) literature has largely emphasized the role of physical appearance and embodiment

of collaborative confederates [23, 36]. As a result, the effect of behavioral and social cues on social facilitation with artificial agents is less discussed. Therefore, the current study adopts the model of Distraction-Conflict Theory [2, 43] to investigate whether human-machine collaboration can also be mediated by attentional conflicts caused by social and behavioral cues. Finally, this study addresses individuals' perception toward machines as another less discussed attribute in former human-machine social facilitation literature. Specifically, we examine whether the perceived human or non-human identity of a cobersational partner, in combination with the social cues of a human- vs non-human conversational style, would potentially influence social facilitation between humans and machines.

3 STUDY 1: REAL BOT VS. REAL HUMAN

3.1 Method

3.1.1 Pilot study. Prior to the formal data collection process, we conducted a pilot study with 28 participants, who were randomly assigned to work with either a robotic chatbot or a friendly chatbot pretending to be a human. Based on our pilot study, we ensured validity of our measurement and refined the study protocol. Using G*Power 3.1, we conducted a power analysis for an independent sample t-test ($\alpha = .05$) to determine the sample size, which is around 30 participants per condition. After piloting, we also pre-registered our study on Open Science Framework (osf.io).

3.1.2 Participants and Procedure. We recruited a sample size of N = 63 (32 in the *Real Bot* condition and 31 in the *Real Human* condition) through Mechanical Turk (MTurk), using the following recruitment criteria: (1) participants located in the United States, (2) HIT approval rate greater than 97%, and (3) more than 500 HITs approved over the worker's lifetime. The mean age of the sample is 37.54 ± 10.86. 58% of the participants identified as males. The experiment was distributed and conducted entirely online via Qualtrics. Participants who consented to participate first completed a pre-survey reporting their self-evaluation of creativity [53] and level of anxiety in group communication [48]. They then worked on a brainstorming task with a conversational partner through online chat. During the brainstorming part of the study, participants in the *Real Bot* condition were directed to an external website where we implemented the chatbot to work on the brainstorming task; participants in the *Real Human* condition would log in to a Google Hangout account with provided credentials to converse with a research assistant. During the brainstorming task, they were asked to come up with creative ideas for water and energy conservation. Afterwards, participants responded to a post-study survey regarding their experiences in the brainstorming task just then and their evaluation toward their conversational partner. We specified in the research consent form that participants were required to take part in the brainstorming task in order to receive compensation, and thus, participants who did not interact at all with their partner nor contribute any ideas during the task were excluded from data analysis. Participants who successfully completed the study were compensated with $7 cash. All protocols were reviewed and approved by the Institutional Review Board (IRB).

3.1.3 Study Materials. Using a consumer chat bot platform, Juji.io, we created *Idea Bot* for the *Real Bot* condition. We used the Juji Studio user interface to design the conversational flow. Additionally, we used Juji IDE to revised the code and customized certain functionalities of the chatbot to accommodate the need for the present research. Specifically, we prevented the chatbot from retrieving Juji's pre-built-in dialogue libraries, so that the chatbot would only follow the same conversational script written for the study. We also included a timer function in the chatbot, so that when the conversation reached 10 minutes, the bot would suggest to wrap up. To simulate the experience of speaking with an robotic confederate, we intentionally made the language used in the conversation more polite and formal. For example, the chatbot would start the conversation with "Hi there! I'm Idea Bot. I'm your AI helper to come up with cool ideas."

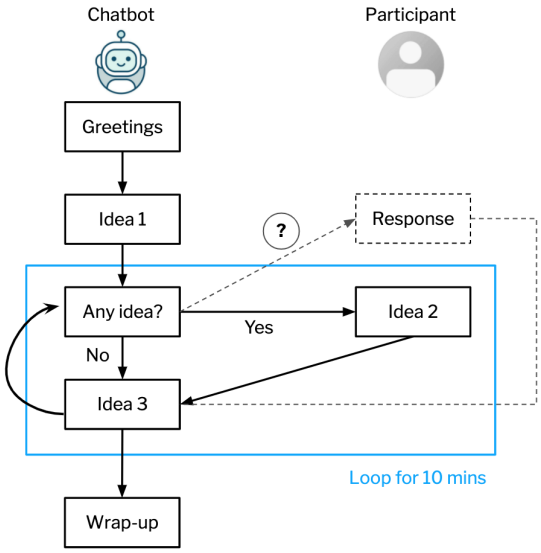
We used a cartoon robot image as the avatar of the chat bot. Unlike most of the modern AI-mediated chatbots, the bot implemented in the present research did not embed any real-time active learning features, since our goal is not to build a human-like chatbot, but to simulate a collaborative experience that elicits less social pressure and is distinguishable from interacting with human confederates. Therefore, we implemented a rather simple chatbot following the same conversational flow when interacting with participants in order to focus on comparing the effect of perceived identity and conversational style on human-machine teamworking outcomes.

The chatbot applied the same conversational flow when interacting with all participants. The chatbot would first greet the participants, introduce itself, and invite participants to start working on the task. Following this, the chatbot would propose the first idea, ask participants' opinions about the idea, and ask whether they had other ideas in mind. If participants did not have an idea to contribute, the chatbot would move on to propose the next idea. If the bot could not understand what participants said, the default response was "Tell me more about that!" It would then wait for participants' response, and move on to propose the next idea. During the 10-minute brainstorming task, the chatbot could potentially contribute as many as ten ideas, which include (1) collecting and reusing rain water, (2) installing smart toilets to water usage when flushing, (3) installing water-saving filters on fountains and faucets, (4) installing small solar panels to support electricity usages in household, (5) installing small wind mills to support energy usages, (6) going "paper-less" and printing as few materials out as possible, (7) bringing your own shopping bags to grocery shopping, (8) bringing your own containers or bottles to restaurants and coffee shops, (9) eating less meat and consuming more vegetarian/plant-based food, (10) carpooling and taking public transportation more often instead of driving your own vehicles. These ten ideas were selected from the top ten ideas that participants most often came up with in a previous study that used the same prompt.

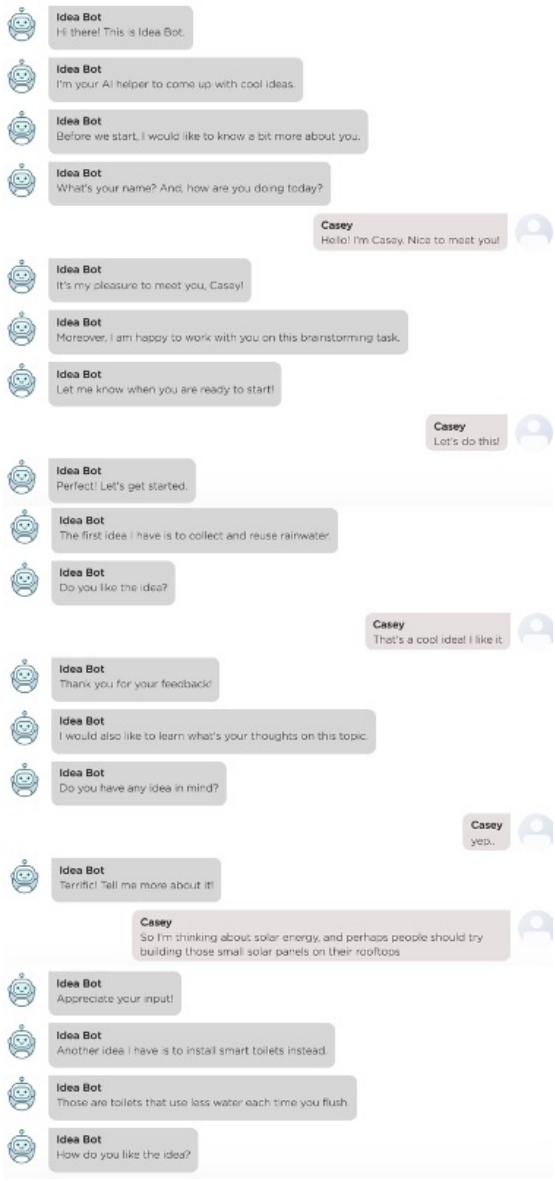
The research assistant also followed a study script to interact with participants, which also started with greetings and self-introduction. The research assistant proposed the same set of ten ideas, but the order of presentation depended on the conversation with participants. The research assistant could also respond appropriately to what participants said.

3.1.4 Measures. Adopting a Triangulation approach [17], we captured participants' idea generation experience and outcomes through a diverse set of variables, including qualitative and quantitative measurement. First of all, we recorded chat transcripts of during all conversations, and we analyzed participants' brainstorming processes using the Linkography method [18, 29]. Secondly, we included three open-ended questions in the post-survey, asking participants to describe their thought processes during the brainstorming task, and to describe their conversational partner. Thirdly, our pre- and post-surveys included various quantitative variables according to categories in our theoretical framework, which we elaborate below. All quantitative variables used in the current research were measured on 7-point Likert scales unless specified otherwise. Table 1 provides descriptive statistics of all quantitative measures.

Dependent Variables: Dependent variables included both researcher-coded and participants' self-reported measures of task performance. **Idea generation outcomes** were coded from participants' chat transcripts, and evaluated on six measures [30, 45]. These include (1) *the number of ideas generated.* (2) *The length of idea description* calculated by word-counts of each idea. (3) *Originality of ideas*, coded on a 3-point scale. An idea's originality was coded as 1 if it is an existing idea, as 2 if participants combined an existing idea with their own new input, and as 3 if it was an completely new idea that has not been commonly practiced for water and energy conservation. (4) *Logic with partner* is coded in a binary fashion, capturing whether an idea is (1) or is not (0) related to what their partner previously proposed. (5) Similarly, *Logic with oneself* is also a binary measure, denoting whether the idea connected to what participants



(a) The conversational flow of Idea Bot (Dashed lines denote cases when the bot cannot understand what participants said).



(b) A screenshot of conversation with Idea Bot

Fig. 3. The conversational flow design of Idea Bot with an example

previously mentioned themselves or not. (6) *Load of information* in ideas is coded on a scale of 3-point scale, measuring the amount of information embedded in the idea.

Self-report measurement of task performance was captured by five items of **creative self-efficacy** (Cronbach's $\alpha = .96$) adopted from Tierney et al.'s literature [53]: (1) "I have confidence that I solved the task creatively." (2) "I had a

Table 1. Descriptive Statistics of the Three Studies

Experiment	Study 1		Study 2		Study 3	
	M	*S.D.*	*M*	*S.D.*	*M*	*S.D.*
Anxiety in group communication	3.87	1.23	4.36	1.35	4.12	1.26
Idea generation outcomes						
Count of ideas	5.31	3.55	4.03	3.06	4.65	3.36
Length of idea	19.33	10.83	14.70	7.94	16.90	9.69
Originality	1.27	.36	5.11	3.07	3.27	2.94
Load of information	1.80	.54	1.46	.61	1.63	.59
Logic with oneself	.25	.20	.26	.20	.24	.23
Logic with partner	.28	.27	.27	.17	.28	.21
Creative self-efficacy	5.06	1.29	5.38	1.01	5.24	1.18
Perception toward partner						
Perceived dominance	2.70	1.66	3.40	2.06	3.89	1.79
Perceived productivity	5.95	1.07	5.90	1.09	6.00	1.03
Perceived creativity	4.77	1.88	5.11	1.34	4.96	1.58
Concerns about partner	3.48	1.83	4.40	1.85	3.93	2.08
Conversational experiences						
Express thoughts	6.38	.87	6.25	.85	6.24	.79
Pause and wait	5.92	1.29	5.78	1.12	5.44	1.69
Pressure to contribute	3.81	2.01	4.28	1.81	4.70	1.89

knack for further developing the ideas of others." (3) I feel that I was good at generating novel ideas during the task." (4) I had a lot of good ideas during the task." (5) "I had good imagination during the task."

Moderating variables: Twelve items, including both positive and negative measures of the level of **anxiety in group communication** (Cronbach's $\alpha = .91$), was Stephen et al.'s literature [48]. Positive measures of group anxiety asked whether participants tend to "follow rather than lead a conversation," "give vague answers and do not often take a stand," "sensitive to criticism," "show nervousness when speaking in front of public," "often say the first thing that comes to mind," "hold back in group conversation." Negative measures surveyed whether participants "express ideas well, speak easily and smoothly," "answer questions with a simple 'yes' or 'no'," "control what gets talked about," "take the initiative, offer suggestions, information, or plans," "dominate others in conversation," and "appear confident" in a group conversation.

Mediating variables: Based on the theoretical framework of the present research, these mediating variables can be broken down to two categories. (1) **Perception toward teamworking partners:** Per our literature review, we measured participants' perceived dominance and productivity of their partner. Given the context of the present experiment (i.e., teamworking on a brainstorming task), we also asked participants to evaluate how creative their partners were. (2) **Conversational experiences** capture whether participants could freely express their thoughts in the conversation, whether they pause and wait for their partners while noting the (...) sign when their partners were inserting messages, and whether they felt pressure to come up with ideas when they were asked to contribute to the task [33, 35].

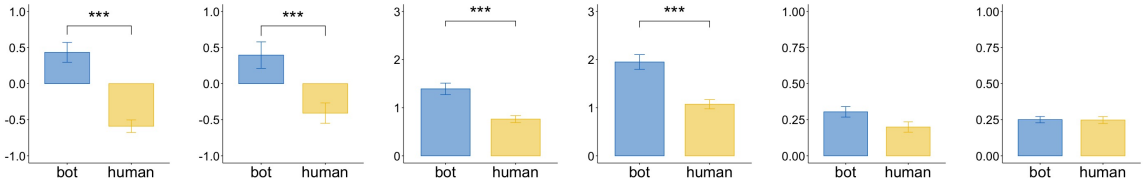
Manipulation check: To ensure the success of our experimental manipulation, we asked participants whether they considered their conversational partners as "absolutely bot" or "absolutely human." Participants responded to this question using a 100% slider scale. We placed this question at the very end of the survey to prevent it from priming participants' suspicion for the real identity of their partner and thus affecting their responses to other questions.

3.1.5 Plans for Data Analysis. We used two-tailed, independent sample t-tests to evaluate the main effect of experimental conditions on dependent variables. To test the significance of moderating effects, we applied general linear models from the lmer package in R. including participants' unique IDs as a random factor. Finally, we performed path analysis to test mediation, using the single-mediator SEM model of the lavaan package in R. For analysis measurement without a fixed scale (e.g., the total count of ideas generated, the length of idea description), we applied its standardized z-scores for analysis. Furthermore, based on convention in statistics literature [51], we excluded data with values outside the range of $M \pm 3.29S.D.$ to mitigate the effect of extreme values.

3.2 Results

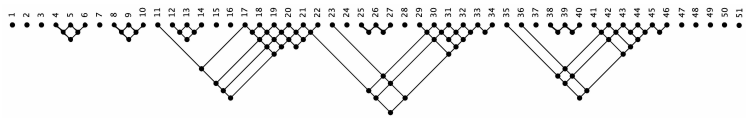
We began our data analysis by assessing whether the experimental manipulation was successful. To perform the manipulation check, we conducted a two-tailed independent sample t-test. Test results showed significant mean difference of partner's realness between the two conditions ($t = -9.11, p < .001$), where participants in the *real human* condition were more likely to rate their conversational partner as a real human confederate ($M = 83.25, S.D. = 4.55$) than participants in the *real bot* condition did ($M = 16.83, S.D. = 5.61$). Following, we adopted the analytic plans as mentioned above to test whether the experimental conditions, moderating variables, and mediating variables produced significant effect. Summary statistics of all test results can be found in Table 2.

Idea generation outcomes: Using two-tailed independent sample t-tests, we first examined whether the quantity of ideas generated by participants differs between the two conditions, using the standardized z-scores for analysis. We found participants in the *real bot* condition contributed significantly more ideas than those in the *real human* condition ($t = 6.24, p = .0003$). Next, we looked at the length of idea description by word count. Again, we performed the t-test with standardized z-scores, excluding extreme values outside the range of ±3.29 [51]. Results showed participants responded with a greater length of idea description in the *real bot* condition ($t = 3.46, p = .001$). When evaluating each idea generated, participants in the *real bot* condition came up with ideas that were rated as more original ($t = 4.47, p = .00003$) and with a greater load of information ($t = 4.83, p = .000009$). In terms of logical connection between ideas, participants in the *real bot* condition were more likely to relate to former ideas of their own ($t = 3.43, p = .001$). However, there were no significant difference in the number of logical connections whether participants built their ideas upon a bot's or a human partner's ($t = 1.67, p = 0.100$). When investigating participants' linkographs, we also see more complex, interrelated patterns of brainstorming process for those in the *real bot* condition, where as patterns in the *real human* condition were relatively sparse.

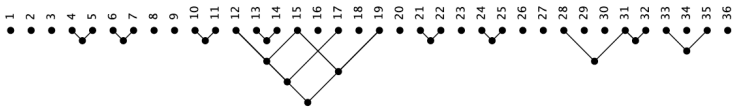


(a) count of idea (b) length of idea (c) originality (d) load of info (e) logic w/ oneself (f) logic w/ partner

Fig. 4. Study 1: Main effect on idea generation outcomes. (The y-axis denotes in each subfigure denotes value of the variable as specified in the subfigure's header. Values shown in subfigure (a) and (b) are standardized z-scores of the variables.)



(a) The Linkograph of the participants who generated the most ideas in the *Real Bot* condition



(b) The Linkograph of the participants who generated the most ideas in the *Real Human* condition

Fig. 5. Study 1: Examples of high-performing participants' Linkographs

Creative self-efficacy: Based on an independent sample t-test, there was no direct main effect of experimental conditions on creative self-efficacy ($t = .73, p = .500$). Through a general linear model, we found participants' self-efficacy was significantly moderated by their anxiety in group communication ($F(3, 59) = 1.64, R^2 = .07, p < .042$). The two-way interaction is further probed by Johnson-Neyman technique to define its statistically significant region. Specifically, we found highly anxious participants (anxiety > 5.04 on a 7-point scale) reported significantly higher creative self-efficacy when brainstorming with a bot, compared to those interacting with a human confederate.

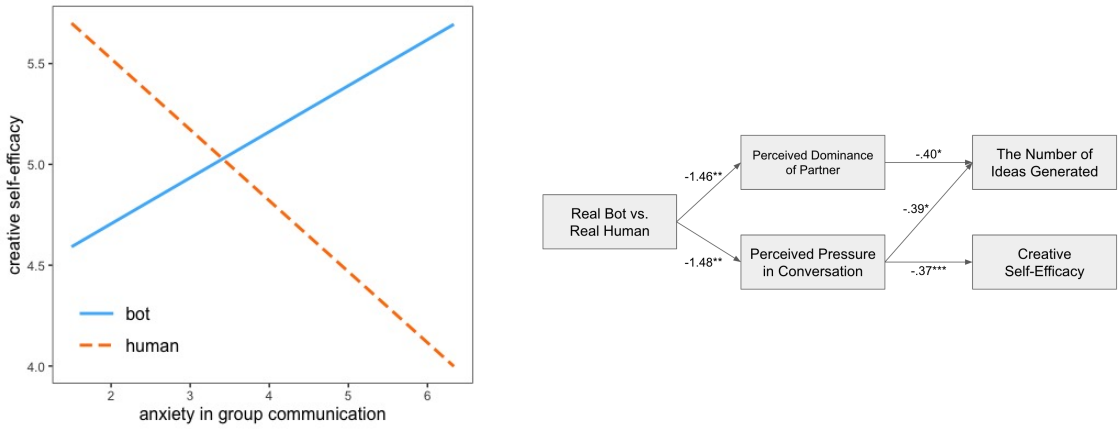


Fig. 6. The moderating effect of anxiety in group (left) and the mediating effect of perceived dominance and pressure (right)

Perception toward teamworking partners: Through independent sample t-tests, we found that participants rated their bot partner as significantly more dominant than the human confederate ($t = 2.82, p = .006$). On top of that, we performed a SEM path analysis using the lavaan package in R. Results suggested the effect of experimental conditions on the total number of ideas generated was significantly mediated by the degree of perceived dominance of their conversational partners ($\beta = 14.30, S.E. = 7.11, p = .021$). More specifically, the finding that participants in the *real bot* condition came up with more ideas can be partially explained by the higher level of dominance they sensed from their chatbot partner. Otherwise, participants in the two conditions viewed their conversational partners as equally

productive ($t = -.34, p = .700$) and creative ($t = -1.55, p = .100$). Additionally, participants were no less concerned about the bot partner's feedback than that of the human confederate ($t = .10, p = .090$).

Conversational experiences: According to results of independent sample t-tests, we found participants in the *real bot* condition felt greater pressure to come up with ideas when their partner asked for contribution ($t = 3.08, p = .003$). We again performed a SEM path analysis, based on which we found participants' perceived pressure in the conversation significantly mediated their creative self-efficacy ($\beta = .40, S.E. = 0.17, p = .021$) and marginally mediated the total number of ideas generated ($\beta = .56, S.E. = 0.40, p = .058$). In both conditions, participants reported they were equally likely to pause and wait for their collaborators while they saw the sign of them inserting messages ($t = 1.47, p = .100$). Moreover, in both conditions, participants reported that they were able to freely express their thoughts ($t = 1.52, p = .100$).

3.3 Discussion

In Study 1, we found participants generated more ideas and with ideas of better quality (i.e., more original, with more detailed descriptions, and with a greater load of information) when they interacted with a chatbot in a brainstorming task. In the *real bot* condition, participants were also more focused on their idea generation process, demonstrating more logical connections among their own ideas. This result aligns with our hypothesis based on the Social Facilitation Theory: because participants would worry less about judgments from their bot partners, they yielded more fruitful outcomes in the task. Additionally, there was a significant interaction with self-reported anxiety, with participants who demonstrated greater anxiety in group communication reporting a higher degree of creative self-efficacy when they interacted with a bot partner. Based on our mediation analyses, the positive effect of teamworking with a bot on idea generation outcomes may be explained by the perceived dominance of conversational partners as well as the amount of pressure to come up with new ideas. However, besides the perceived identity of collaborators, the different conversing styles of a human and a bot can also affect task outcomes. In particular, the fast-paced conversing style and the fewer social exchanges in dialogues with a chatbot can also account for the greater number of ideas generated. Therefore, to further understand how the perceived identity and conversing style of conversational partner may lead to distinct teamworking outcomes, we conducted a follow-up study using a human research assistant confederate for both the "human" and "bot" conditions as described below in Study 2.

4 STUDY 2: FAKE BOT VS. REAL HUMAN

Based on findings in Study 1, we further asked whether the positive teamworking outcomes of idea generation were due to the perceived identity or the conversational style of collaborators. Therefore, in Study 2, we controlled the conversing styles by having a human confederate interact in a friendly human tone in both conditions. With this experiment design, we intended to answered the following research questions:

RQ1a: Given a humanoid conversing style, how does the perceived identity of a non-human confederate influence the teamworking outcomes of idea generation?

RQ2a: Given a humanoid conversing style, how does the level of anxiety in group communication moderate teamworking and idea generation with a non-human confederate?

RQ3a: Given a humanoid conversing style, how does the perception toward teamworking partners and conversational experiences mediate teamworking and idea generation with non-human confederates?

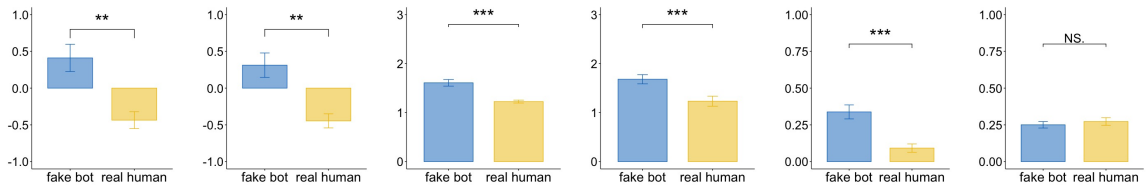
4.1 Method

To understand the effect of perceived identity of collaborative confederate on teamworking outcomes and experiences, we conducted a follow-up study by replicating the materials and procedures of Study 1, with the exception that all participants in Study 2 worked with a human confederate. The purpose of this study design was to control the conversing style of the teamworking partner, leaving perceived identity as the sole manipulated variable. Therefore, all participants logged in to a pre-registered Google Hangout account and chatted with the same human partner. Participants in the *real human* condition were informed that their partner was a research assistant, and participants in the *fake bot* condition were told that they were collaborating with a chatbot. Again, participants were recruited through MTurk and were randomly assigned to one of the two conditions. All measures in Study 2 included the same measures as in Study 1. However, we implemented a more rigorous approach to manipulation check. Other than comparing mean difference from the slider scale question ("absolutely bot" vs. "absolutely human") at the end of the survey, we also referred to the two open-ended responses, where participants described their conversational partners and teamworking processes. Accordingly, two research assistants rated whether participants viewed their partners as a bot or a human. Participants who were in the *real human* condition but firmly asserted their partner was a bot were excluded from data analysis, and vice versa. After filtering out disqualified data, data from N = 68 participants were adopted for analysis (gender: m = 42, f = 25, age: $M = 37.8, S.D. = 12.2$).

4.2 Results

We again started by performing a two-tailed independent sample t-test for the manipulation check. The test result suggested a successful manipulation in the experiment ($t = 3.46, p = .0009$), where participants in the *real human* condition were more likely to believe their conversational partner was, indeed, a human ($M = 73.21, S.D. = 5.22$) than those in the *fake bot* condition ($M = 47.65, S.D. = 5.21$).

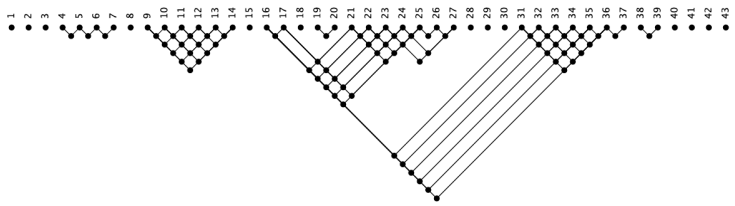
Idea generation outcomes: Aligning with Study 1's findings, independent sample t-tests also revealed significant mean differences in the quantity of ideas ($t = 3.83, p = .0003$), length of idea description ($t = 3.88, p = .0002$), originality of idea ($t = 2.75, p = .008$), load of information ($t = 2.93, p = .002$), logical connection with oneself ($t = 4.48, p = .00004$). In particular, more ideas were generated and with higher quality in the *fake bot* condition. Again, there was no significant difference in the likelihood of building on partners' ideas between the two conditions ($t = .66$, p is n.s.).



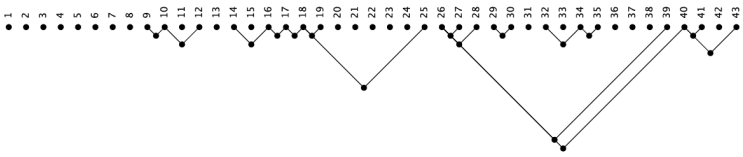
(a) count of idea (b) length of idea (c) originality (d) load of info (e) logic w/ oneself (f) logic w/ partner

Fig. 7. Study 2: Main effect on idea generation outcomes. (The y-axis denotes in each subfigure denotes value of the variable as specified in the subfigure's header. Values shown in subfigure (a) and (b) are standardized z-scores of the variables.)

Creative self-efficacy: Once again, we did not observe a significant main effect of the experimental conditions on participants' creative self-efficacy ($t = .53, p = .60$). However, in Study 2, we also no longer observed a significant moderating effect of participants' anxiety in group communication ($F(3, 59) = 0.43, R^2 = .02, p = .47$). We discuss possible explanations for this finding below.



(a) The Linkograph of the participants who generated the most ideas in the *Fake Bot* condition



(b) The Linkograph of the participants who generated the most ideas in the *Real Human* condition

Fig. 8. Study 2: Examples of high-performing participants' Linkographs

Perception toward conversational partners: Compared to Study 1, participants evaluated their conversational partner quite differently than in in Study 2. To begin with, the difference in perceived dominance was only marginally significant ($t = 1.91, p = .06$), where participants viewed the *real human* partner as slightly more dominating. As a result, when we tested the mediating effect of partner's dominance, the result in Study 2 was no longer significant. Furthermore, participants reported the human confederate as more creative ($t = 2.84, p = .006$) and more productive ($t = 2.63, p = .010$) than the supposed "bot" confederate. On top of that, participants also expressed more concern about their human partner's feedback ($t = 2.28, p = .030$).

Conversational experiences: Similar to Study 1, self-report measurement did not reveal any significant differences in whether participants could freely express themselves ($t = .21, p = .80$) and whether they paused and waited for their teammates ($t = .37, p = .70$ is n.s.). Distinct from Study 1, participants reported greater pressure to come up with new ideas in the *real human* condition. However, the between-group difference was less significant than that in Study 1 ($t = 2.17, p = .03$). As a result, the mediating effect of perceived pressure on creative self-efficacy and the total number of ideas generated was not replicated in Study 2.

4.3 Discussion

Despite applying the same conversing style, Study 2 replicated the main effect of teammates' perceived identity on positive idea generation, as observed in Study 1. However, the moderating effect of group anxiety and the mediating effect of conversational experiences and perception toward partners were no longer significant. Based on this results, we asked whether the two-way interaction and mediation in Study 1 can be attributed specifically to the robotic conversing style of the bot partner. To tackle this inquiry, we conducted Study 3 as a final follow-up study.

5 STUDY 3: REAL BOT VS. FAKE HUMAN

As mentioned above, the motivation to implement Study 3 was not only to confirm whether the main effect of experimental conditions in Study 1 was dependent on conversational partners' conversing style, but also, to investigate

whether the moderating and mediating effect on creative self-efficacy is unique to experiencing a robotic conversing style. Therefore, we proposed additional research questions for Study 3:

RQ1b: Given a robotic conversing style, how does the perceived identity of a non-human confederate influence the teamworking outcomes of idea generation?

RQ2b: Given a robotic conversing style, how does the level of anxiety in group communication moderate teamworking and idea generation with a non-human confederate?

RQ3b: Given a robotic conversing style, how does the perception toward teamworking partners and conversational experiences mediate teamworking and idea generation with non-human confederates?

5.1 Method

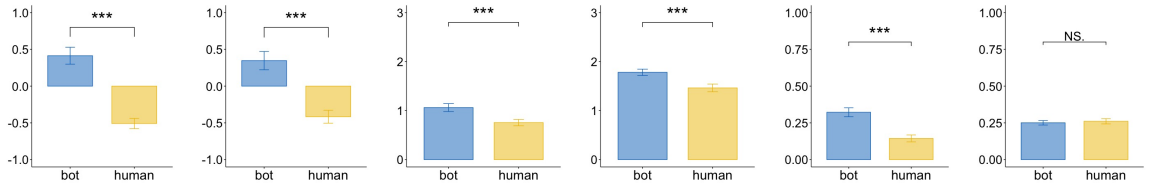
To further understand the effect of conversing style of teamworking partners on participants' creative self-efficacy, as well as to confirm the positive idea generation outcomes when working with a chatbot partner, we conducted a third study. In Study 3, we replicated the materials and procedures of Study 1 and 2, with one exception. In this study, participants all worked with a chatbot, but those in the *real bot* condition were informed with the true identity of their collaborator, while those in the *fake human* condition were told that the partner is a human. Participants were again recruited from MTurk. Again, since we were concerned that participants in the *fake human* condition of Study 3 would be more likely to question the actual identity of their conversational partner, we applied both quantitative (i.e., responses to the "absolutely bot" vs. "absolutely human" slider question) as well as open-ended responses for data screening. Two research assistants again involved in the review process to filter out participants who were in the *fake human* condition but strongly believed that their partner was a bot. After excluding disqualified data, the sample size of Study 3 is N = 54 (gender: m = 32, f = 21; age: $M = 37.5, S.D. = 11$) There were 28 participants in the *real bot* condition and 26 participants in the *fake human* condition.

5.2 Results

We again began our analysis by performing a manipulation check with an independent sample t-test. The test result showed a significant difference between the group mean of participants perception toward their partners' identity ($t3.87, p = .0003$). Specifically, those in the *real bot* condition were significantly more likely to consider their partners as a bot ($M = 15.88, S.D. = 11.24$), while those in the *fake human* condition were more likely to believe that their partners were a human ($M = 71.69, S.D. = 15.10$).

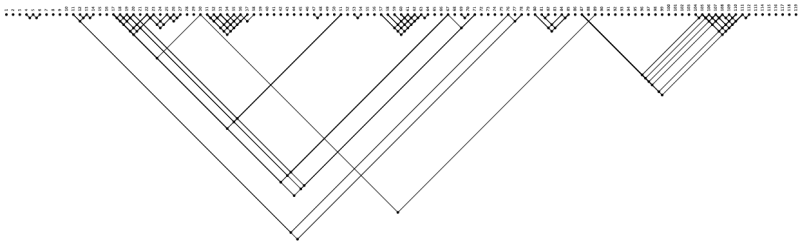
Idea generation outcomes: Just as in Study 1 and Study 2, we observed more fruitful outcomes of idea generation in the *real bot* condition. With independent sample t-tests, we found that when participants were informed that they would be working with a bot partner, they produced significantly more ideas ($t = 3.83, p < .001$). The ideas generated with bot partner contained more detailed idea description ($t = 3.88, p < .001$), greater originality ($t = 2.75, p < .01$), more information ($t = 2.93, p < .01$), and more logical connection with oneself ($t = 4.48, p < .001$). Again, there was no significant difference in the likelihood of building on partners' ideas between the two condition ($t = .66, p$ is n.s.).

Creative self-efficacy: Aligned with Study 1 and 2, there is no significant difference in creative self-efficacy between the two conditions ($t = .79, p = .50$) according to the test result of independent sample t-test. Nonetheless, we observed a marginally significant two-way interaction between the perceived identity of teamworking partners and the level of anxiety in group communication ($F(3, 50) = 1.01, R^2 = .06, p < .1$). The interaction effect was further probed by Johnson-Neyman technique, where the test results suggest a statistical significant region at the moderator's value (i.e.,



(a) count of idea (b) length of idea (c) originality (d) load of info (e) logic w/ oneself (f) logic w/ partner

Fig. 9. Study 3: Main effect on idea generation outcomes. (The y-axis denotes in each subfigure denotes value of the variable as specified in the subfigure's header. Values shown in subfigure (a) and (b) are standardized z-scores of the variables.



(a) The Linkograph of the participants who generated the most ideas in the *Real Bot* condition

(b) The Linkograph of the participants who generated the most ideas in the *Fake Human* condition

Fig. 10. Study 3: Examples of high-performing participants' Linkographs

the level of anxiety in group communication) greater than 5.50 on a 7-point scale. That is, participants who were highly anxious in teams reported significantly greater creative self-efficacy when they were working with a bot.

Perception toward conversational partners: Similar to Study 2, given the same conversing style, participants considered their human partner as more dominating, while the mean difference between the two conditions in Study 3 was more significant ($t = 3.30, p < .01$). On top of this, we also performed a path analysis and found perceived dominance of partners had a significant mediating effect on the number of ideas generated ($\beta = -1.98, S.E. = .99, p < .05$). However, the direction of the mediation is distinct from Study 1. More specifically, participants who perceived their partners as more dominating tended to contribute fewer ideas. Thus the finding that participants in the *fake human* condition yielded a smaller number of total ideas could be partially due to the fact that their partner was perceived as more dominating.

Conversational experiences: Using independent t-tests, we found several main effects of the experimental conditions on conversational experiences. To begin with, participants in the *real bot* condition suggested they could express their thoughts in the conversation more freely than those in the *fake human* condition did ($t = 2.60, p < .01$). Participants were also more likely to pause and wait for their bot partner when it was producing a new message ($t = 2.27, p < .05$). However, there was no significant difference in the amount of pressure to come up with new ideas during the task ($t = .19$, p is n.s.).

Table 2. Summary Statistics of the Three Studies

Note: B: Bot condition; H: Human condition; X: Main effect of experimental conditions; X*W: Moderating effect of the level of anxiety in group communication; M: Mediating effect of perception toward partners or conversational experiences

Experiment	Study 1					Study 2					Study 3				
Condition	Real Bot	Real Human				Fake Bot	Real Human				Real Bot	Fake Human			
Perceived identity	Bot	Human	X	X*W	M	Bot	Human	X	X*W	M	Bot	Human	X	X*W	M
Conversational style	Bot	Human				Human	Human				Bot	Bot			
Idea generation															
Count of ideas	6.85	3.62	B>H			5.29	2.70	B>H			6.04	3.19	B>H		
	(3.72)	(1.76)				(3.36)	(1.03)				(3.16)	(1.39)			
Length of ideas	20.10	17.44	B>H			16.20	13.01	B>H			18.06	15.06	B>H		
	(9.38)	(12.27)				(7.63)	(8.06)				(8.53)	(10.82)			
Originality	1.39	.76	B>H			1.61	1.22	B>H			1.06	.76	B>H		
	(.67)	(.40)				(.57)	(.35)				(.66)	(.51)			
Info load	1.95	1.07	B>H			1.68	1.32	B>H			1.78	1.46	B>H		
	(.86)	(.53)				(.55)	(.58)				(.52)	(.63)			
Logic w/ self	.31	.11	B>H			.34	.16	B>H			.32	.14	B>H		
	(.20)	(.09)				(.57)	(.09)				(.25)	(.18)			
Logic w/ partner	.25	.24	B=H			.25	.27	B=H			.25	.26	B=H		
	(.12)	(.13)				(.13)	(.15)				(.13)	(.14)			
Creative self-efficacy	5.17	4.94	B=H	sig.		5.45	5.32	B=H	n.s.		5.35	5.14	B=H	sig.	
	(1.25)	(1.34)				(.97)	(1.07)				(.81)	(1.44)			
Partner															
Dominance	3.25	2.13	B>H		sig.	2.94	3.38	H>B		n.s.	3.18	4.65	H>B		sig.
	(1.80)	(1.31)				(1.94)	(2.10)				(1.44)	(1.81)			
Creativity	5.34	5.10	B=H		n.s.	5.07	5.58	B<H		n.s.	5.04	5.21	B>H		n.s.
	(1.31)	(1.25)				(1.06)	(1.25)				(1.28)	(1.26)			
Productivity	5.91	6.00	B=H		n.s.	5.57	6.24	H>B		n.s.	6.15	5.86	B=H		n.s.
	(1.12)	(1.03)				(1.27)	(.75)				(.68)	(1.27)			
Concern	3.50	3.45	B=H		n.s.	3.91	4.91	H>B		n.s.	3.81	4.04	B=H		n.s.
	(2.06)	(1.59)				(1.99)	(1.57)				(1.88)	(2.28)			
Conv. Experiences															
Expressiveness	6.22	6.55	B=H		n.s.	6.23	6.27	B=H		n.s.	5.96	6.50	B>H		n.s.
	(.91)	(.81)				(.91)	(.80)				(.77)	(.74)			
Pressure	4.53	3.06	B>H		sig.	3.83	4.76	H>B		n.s.	4.65	4.75	B=H		n.s.
	(1.95)	(1.82)				(1.67)	(1.85)				(1.65)	(2.12)			
Pauses	5.69	6.16	B=H		n.s.	5.83	5.73	B=H		n.s.	4.92	5.93	B>H		n.s.
	(1.42)	(1.10)				(.95)	(1.28)				(1.90)	(1.33)			

5.3 Discussion

Once again, Study 3 confirmed the effect of the perceived identity of teamworking partners on idea generation outcomes. Agreeing with the former two studies, participants contributed more ideas with greater quality when they were informed that they would be interacting with a bot partner. Additionally, highly anxious participants demonstrated greater creative self-efficacy when they perceived their partner as a non-human agent. Though the interaction effect in Study 3 was less significant than that in Study 1, we suspect this result can largely be influenced by participants in the *fake human* condition being more suspicious of their partners' identity. That is, unlike those in the *real human* condition of Study 1, participants interacting with a bot pretending to be a human may not have firmly believed that the real identity of their partner is indeed a human being. The significant mediating effect of perceived dominance of a teamworking partner was again observed in Study 3. However, while participants perceived their human partners as more dominating, this resulted in fewer contributions. The results can be relevant to the social loafing tendency in teams. Specifically, former research in teamworking suggested that, with a overly dominating opinion leader in a group, other teammates may become less motivated to contribute [3, 19, 21, 38].

6 GENERAL DISCUSSION

6.1 The effect of perceived identity

Throughout a series of three studies, we repeatedly found participants generated more ideas and contributed ideas with better quality when they perceived their collaborator as a non-human confederate. We consider the findings can be explained through various perspectives. The first, based on the Social Facilitation Theory [1, 59] and the Distraction-Conflict Framework [2, 43], we propose that working with a non-human agent may reduce the distraction of concerns about one's teammates. This notion is not only supported by the fact that highly anxious participants attained greater creative self-efficacy when they were told they were interacting with a bot partner, but also, we saw participants in the bot conditions were more likely to connect the dots and relate to their previous ideas during the task. This suggested that they were concentrating on their own brainstorming processes, which can be another positive factor leading to more fruitful outcomes for idea generation.

Secondly, though we did not directly measure this as one of our quantitative variables, teamworking with a chatbot may be more likely to be perceived as a fun, novel experience. Teamwork with a bot is relatively novel for the general public, compared to other usages of chatbots (e.g., customer services). In fact, based on participants' open-ended responses, those in the bot conditions commonly described their partner or experiences as "cool," "fun," or "more interesting than expected." Some participants may even have viewed the teamworking task to have a gamification component. For instance, a small number of participants specified in their open-ended responses that they attempted to compete and to come up with more ideas than the bot did. Under the Distraction-Conflict Framework [2, 43], perceived novelty and interest of the conversational experience may have increased the degree of alertness and engagement while participants were working on the task.

Other literature in HRI suggests that humans may adopt distinct cognitive states when interacting with a robotic agent versus a human [8, 40]. Though to date, there is limited understanding of how such unique mental states when interacting with robots may facilitate creativity in teams, one particular aspect may be highly relevant to the present research. That is, former studies have found individuals tend to expect repeated patterns in conversational and interactive flows of robots [14]. In our studies, both the chatbot and the research assistant followed a structural framework to

conduct the conversation. Hence, when informed that their partner is a bot, participants may not spend time trying to engage in spontaneous, social interaction, and once again, stay focused on the process of generating ideas per se.

When it comes to expectation for autonomous agents, our findings also reveal insights regarding perceptions toward bot partners. In recent robot and AI-mediated communication literature [25, 49, 55], studies have found the presence of bots in group settings can serve as a social buffer when conflicts and negative affects arose. Specifically, by partially attributing inappropriateness to artificial agents, participants blamed their human teammates less in tense circumstances. Therefore, negative perceptions toward artificial confederates in these team settings can, in fact, be advantageous. We propose that a similar effects may explain participants' reaction to the perception of dominance from their bot partner. That is, though being a dominating teammate is typically associated with an inferior impression among *humans*, the negative perception of a dominating, pressing *bot* could possibly lead them to more productive outcomes, and eventually greater self-efficacy in the degree of creativity demonstrated in their task performance.

6.2 The effect of conversing styles

Through these three studies, we found that certain outcomes of interacting with a non-human agent were dependent on the conversing styles applied by the confederate. We found that the moderating effect of experimental conditions and the level of anxiety in group communication was significant in Study 1 but not in Study 2. Previous research in HRI and anthropomorphism suggested humans have the tendency to view other subjects as humans when they perceive humanoid cues (i.e., a human conversational style as applied in Study 2) [15, 61]. This effect may be more prominent among highly anxious participants, since their cognitive capacity had already been predominantly occupied by engaging in a team setting. With less cognitive control, highly anxious participants may tend to treat their human-acting *fake bot* partner more like a human partner. Therefore, the positive effect of perceived bot identity did not make an impact on these anxious participants in Study 2.

Next, when we compared the mediation effect of perceived dominance of partners in Study 1 and Study 3, we can see how different conversing styles influence participants' perception toward their partner, and thus affecting their teamworking and brainstorming outcomes. That is, when labeled with a human identity, teamworking partners were perceived as more dominating when they applied a robotic conversing style. Interestingly, participants' reaction to dominating partners were also different, depending on whether they were interacting with a human or a non-human agent. Specifically, when participants believed their partner was a human, they relied on their partners for the teamworking task and demonstrated a tendency of social loafing. However, when they believed they were interacting with a dominating bot, the perceived dominance drove participants to be more involved in the task. Again, based on the Distraction-Conflict framework [2, 43], the unique experience of interacting with a bot teammate may again increase the alertness, attention, and engagement of participants.

6.3 Design Implications

Findings of the present research, specifically those relevant to the conversing style of non-human confederates, offer rich insights for design implications of future chatbots. Particularly, we would like to emphasize the outcomes of our current study does not set a goal to design autonomous agents that work "better" than human teammates, but ones that are distinguishable from humans. In fact, based on our findings, signaling social-behavioral cues that are more robotic can be beneficial to solving a creative teamworking task. Additionally, an artificial identity embedded in an active team player can also bring benefits to teamworking, such as perceived novelty, interest, and positive competition, which may all contribute to greater engagement in a collaborative experience.

On top of that, we suggest future design of artificial agents can adopt a greater degree of user control over bots' conversational styles to facilitate teamworking and idea generation. Specifically, when we take into account individual differences in the level of anxiety in group communication, a robotic-sounding or a human-like conversational style may benefit different groups of participants respectively. Specifically, anthropomorphic features (e.g., physical appearances, avatar images, facial expressions, emotional responses) as well as conversing cues (e.g., pace, formal (or informal) languages, usages of emojis, paraphrasing, and pausing time to wait for users) can be promising areas for user-enabled customization.

Similar design implications may also be insightful for designing autonomous agents for educational purposes, though further research is required. In fact, the brainstorming task used in the current study simulates in-class activities that are commonly applied for learning. The task not only asked participants to generate creative ideas for a specified topic, but the brainstorming process per se also requires them to connect and apply existing domain knowledge in the topic area [12]. To encourage less confident students to speak up and to express their ideas more openly, designers and engineers of bots for classrooms may refer to the positive effect of robotic-signaling features on participants who demonstrate anxiety in groups.

6.4 Limitations and Future Research

There are, of course, a number of limitations in the present research. To begin with, due to the current COVID-19 pandemic, we conducted the study entirely online, instead of as a standard lab study, to avoid in-person contact with participants. Therefore, unanticipated incidents may have occurred on participants' side without our observation. For example, differences in Internet speed and connectivity may also result in different perceptions of the conversational partners' pace and responsiveness, which is a confound out of our control. To implement the study online, we also chose to conduct the brainstorming task over text, which may not be the most common form of idea generation, compared to other formats such as face-to-face communication and video conferencing. Therefore, a possibility for future research is to test the effect of perceived identity and conversational styles of artificial agents on different platforms that allow for more naturalistic conversation.

On top of that, we also acknowledge the fact that both the bot and the human teammates followed a semi-structured script during the task, which is an atypical brainstorming scenario. This methodological decision provided us greater control over the conversation but was less naturalistic. This decision was primarily motivated by our intention to empirically test the effect of two specified variables (i.e., perceived identity and conversational style). In this regard, potential follow-up studies can examine teamworking outcomes by randomly pairing two participants into teams and either inform them with the real (i.e., human) or fake (i.e., bot) identity of their partners. Additionally, examining ideation outcomes when participants work alone on the brainstorming task can also introduce a "true control" to the research design framework.

Next, we address our perspectives regarding the brainstorming task in the present research. This task has been repeatedly used in former HCI studies [50], allowing us to sample typical participant responses, and this task also allowed a simple transition to an online study. However, the task can only partially capture the complexity and dynamics of human creativity. Specifically, compared to other less structural, free-flowing tasks in existing creativity research, such as composing a drawing or a piece of writing, the task does require more convergent and inferential thinking [30, 45]. Hence, another potential avenues for future research is to examine whether the effect of perceived identity and conversational styles of artificial agents influences other creative tasks in a similar way.

Last but not least, we propose adopting features of user-enabled customization as a critical component in the above-mentioned design implications. Therefore, investigating potential moderating or mediating effect of these customization affordances is an important area of future research.

7 CONCLUSION

The present research conducted a series of three studies to investigate the outcomes of idea generation with artificial agents as active team players in a dyadic conversation. Based on a theoretical framework grounded in the Social Facilitation Theory and the Distraction-Conflict Framework, we systematically compared the effect of perceived identity and conversational style of bot vs. human partners. When interacting with confederates that are perceived as a machine-mediated bot, participants consistently produced more ideas, and ideas with higher quality, across the three studies. Additionally, participants' creative self-efficacy can be moderated by individual differences in team settings, where highly anxious individuals demonstrated greater efficacy in task performance when interacting with a robotic-sounding bot. Additionally, whether the pressing, dominating characteristics of a bot partner improved participants' task performance may depend on robotic social-behavioral cues in an agent's conversational style. Overall, the present research supports the promising role of autonomous agents as active teammates, and points to ways in which leveraging the distinction between artificial and humanoid cues can drive positive outcomes.

REFERENCES

[1] Floyd H Allport. 1920. The influence of the group upon association and thought. *Journal of experimental psychology* 3, 3 (1920), 159.

[2] Robert S Baron, Danny Moore, and Glenn S Sanders. 1978. Distraction as a source of drive in social facilitation research. *Journal of personality and social psychology* 36, 8 (1978), 816.

[3] Myriam N Bechtoldt, Carsten KW De Dreu, Bernard A Nijstad, and Hoon-Seok Choi. 2010. Motivated information processing, social tuning, and group creativity. *Journal of personality and social psychology* 99, 4 (2010), 622.

[4] Cindy L Bethel, Merijn Bruijnes, Malte Jung, Christoforos Mavrogiannis, Simon Parsons, Catherine Pelachaud, Rui Prada, Laurel Riek, Sarah Strohkorb Sebo, Julie Shah, et al. 2020. 4.4 Working Group on Social Cognition for Robots and Virtual Agents. *Dagstuhl Reports, Vol. 9, Issue 10 ISSN 2192-5283* (2020), 21.

[5] Jim Blascovich, Wendy Berry Mendes, Sarah B Hunter, and Kristen Salomon. 1999. Social" facilitation" as challenge and threat. *Journal of personality and social psychology* 77, 1 (1999), 68.

[6] Charles F Bond. 1982. Social facilitation: A self-presentational view. *Journal of Personality and Social Psychology* 42, 6 (1982), 1042.

[7] Robert M Bray and Roger Sugarman. 1980. Social facilitation among interacting groups: Evidence for the evaluation-apprehension hypothesis. *Personality and Social Psychology Bulletin* 6, 1 (1980), 137–142.

[8] Cynthia Breazeal. 2003. Emotion and sociable humanoid robots. *International journal of human-computer studies* 59, 1-2 (2003), 119–155.

[9] Cynthia L Breazeal. 2004. *Designing sociable robots*. MIT press.

[10] Charles S Carver and Michael F Scheier. 1981. The self-attention-induced feedback loop and social facilitation. *Journal of Experimental Social Psychology* 17, 6 (1981), 545–568.

[11] Arturo Cruz-Maya, François Ferland, and Adriana Tapus. 2015. Social facilitation in a game-like human-robot interaction using synthesized emotions and episodic memory. In *International Conference on Social Robotics*. Springer, 164–173.

[12] A Dennis and M Williams. 2003. Electronic brainstorming. *Group creativity: Innovation through collaboration* (2003), 160–178.

[13] Elizabeth Duffy. 1962. Activation and behavior. (1962).

[14] Friederike Eyssel, Dieta Kuchenbrandt, and Simon Bobinger. 2011. Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism. In *Proceedings of the 6th international conference on Human-robot interaction*. 61–68.

[15] Julia Fink. 2012. Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *International Conference on Social Robotics*. Springer, 199–208.

[16] Russell G Geen. 1983. Evaluation apprehension and the social facilitation/inhibition of learning. *Motivation and Emotion* 7, 2 (1983), 203–212.

[17] Barney G Glaser and Anselm L Strauss. 2017. *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

[18] Gabriela Goldschmidt. 2014. *Linkography: unfolding the design process*. Mit Press.

[19] Jack A Goncalo and Barry M Staw. 2006. Individualism–collectivism and group creativity. *Organizational behavior and human decision processes* 100, 1 (2006), 96–109.

[20] William B Gudykunst. 1993. Toward a theory of effective interpersonal and intergroup communication: an anxiety/uncertainty management (AUM) perspective. (1993).

[21] Sarah Harvey. 2014. Creative synthesis: Exploring the process of extraordinary group creativity. *Academy of management review* 39, 3 (2014), 324–343.

[22] Beth A Hennessey. 2003. Is the social psychology of creativity really social. *Group creativity* (2003), 181–201.

[23] Nicholas Hertz and Eva Wiese. 2017. Social facilitation with non-human agents: possible or not?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61. SAGE Publications Sage CA: Los Angeles, CA, 222–225.

[24] Jess Hohenstein and Malte Jung. 2018. AI-supported messaging: An investigation of human-human text conversation with AI support. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.

[25] Jess Hohenstein and Malte Jung. 2020. AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust. *Computers in Human Behavior* 106 (2020), 106190.

[26] Crystal L Hoyt, Jim Blascovich, and Kimberly R Swinth. 2003. Social inhibition in immersive virtual environments. *Presence: Teleoperators & Virtual Environments* 12, 2 (2003), 183–195.

[27] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T Hancock, and Mor Naaman. 2019. AI-mediated communication: How the perception that profile text was written by AI affects trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[28] Malte F Jung, Nikolas Martelaro, and Pamela J Hinds. 2015. Using robots to moderate team conflict: the case of repairing violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. 229–236.

[29] Jeff WT Kan and John S Gero. 2008. Acquiring information from linkography in protocol studies of designing. *Design studies* 29, 4 (2008), 315–337.

[30] James C Kaufman and Robert J Sternberg. 2010. *The Cambridge handbook of creativity*. Cambridge University Press.

[31] Sara Kiesler, Aaron Powers, Susan R Fussell, and Cristen Torrey. 2008. Anthropomorphic interactions with a robot and robot–like agent. *Social Cognition* 26, 2 (2008), 169–181.

[32] Christos Kyrlitsias and Despina Michael. 2016. Influence by others' opinions: social pressure from agents in immersive virtual environments. In *2016 IEEE Virtual Reality (VR)*. IEEE, 213–214.

[33] Elizabeth Wolef Morrison and Frances J Milliken. 2000. Organizational silence: A barrier to change and development in a pluralistic world. *Academy of Management review* 25, 4 (2000), 706–725.

[34] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 72–78.

[35] C Nemeth and Brendan Nemeth-Brown. 2003. Better than individuals. *Group creativity: Innovation through collaboration* 4 (2003), 63–84.

[36] Sung Park and Richard Catrambone. 2007. Social facilitation effects of virtual humans. *Human factors* 49, 6 (2007), 1054–1060.

[37] Paul B. Paulus and Vincent R. Brown. 2003. *Enhancing ideational creativity in groups: Lessons from research on brainstorming*. Oxford University Press, Paulus, Paul B., Department of Psychology, University of Texas at Arlington, Box 19528, Arlington, TX, US, 76019, 110 – 136. http://encompass.library.cornell.edu.proxy.library.cornell.edu/cgi-bin/checkIP.cgi?access=gateway_standard%26url=https://search-ebscohost-com.proxy.library.cornell.edu/login.aspx?direct=true&db=psyh&AN=2003-88061-005&site=eds-live&scope=site

[38] Paul B Paulus and Bernard A Nijstad. 2003. *Group creativity: Innovation through collaboration*. Oxford University Press.

[39] Tamyra Pierce. 2009. Social anxiety and technology: Face-to-face communication versus technological communication among teens. *Computers in Human Behavior* 25, 6 (2009), 1367–1372.

[40] Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey. 2007. Comparing a computer agent with a humanoid robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*. 145–152.

[41] Raoul Rickenberg and Byron Reeves. 2000. The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 49–56.

[42] Nina Riether, Frank Hegel, Britta Wrede, and Gernot Horstmann. 2012. Social facilitation with social robots?. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 41–47.

[43] Glenn S Sanders, Robert Steven Baron, and Danny L Moore. 1978. Distraction and social comparison as mediators of social facilitation effects. *Journal of Experimental Social Psychology* 14, 3 (1978), 291–303.

[44] Mohanbir Sawhney, Gianmario Verona, and Emanuela Prandelli. 2005. Collaborating to create: The Internet as a platform for customer engagement in product innovation. *Journal of interactive marketing* 19, 4 (2005), 4–17.

[45] R Keith Sawyer. 2011. *Explaining creativity: The science of human innovation*. Oxford university press.

[46] Kenneth W Spence, IE Farber, and HH McFann. 1956. The relation of anxiety (drive) level to performance in competitional and non-competitional paired-associates learning. *Journal of Experimental Psychology* 52, 5 (1956), 296.

[47] Garold Stasser and Zachary Birchmeier. 2003. Group creativity and collective choice. *Group creativity: Innovation through collaboration* (2003), 85–109.

[48] Timothy D Stephen and Teresa M Harrison. 1986. Assessing communication style: A new measure. *American Journal of Family Therapy* 14, 3 (1986), 213–234.

[49] Brett Stoll, Malte F Jung, and Susan R Fussell. 2018. Keeping it light: perceptions of humor styles in robot-mediated conflict. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 247–248.

[50] Yilu Sun, Omar Shaikh, and Andrea Stevenson Won. 2019. Nonverbal synchrony in virtual reality. *PloS one* 14, 9 (2019), e0221803.

[51] Barbara G Tabachnick, Linda S Fidell, and Jodie B Ullman. 2007. *Using multivariate statistics*. Vol. 5. Pearson Boston, MA.

[52] Susan L Thomas, Linda J Skitka, Stacy Christen, and Mark Jurgena. 2002. Social facilitation and impression formation. *Basic and applied social psychology* 24, 1 (2002), 67–70.

[53] Pamela Tierney and Steven M Farmer. 2002. Creative self-efficacy: Its potential antecedents and relationship to creative performance. *Academy of Management journal* 45, 6 (2002), 1137–1148.

[54] Olivier Toubia. 2006. Idea generation, creativity, and incentives. *Marketing Science* 25, 5 (2006), 411–425.

[55] Margaret L Traeger, Sarah Strohkorb Sebo, Malte Jung, Brian Scassellati, and Nicholas A Christakis. 2020. Vulnerable robots positively shape human conversational dynamics in a human–robot team. *Proceedings of the National Academy of Sciences* 117, 12 (2020), 6370–6375.

[56] Lee Edward Travis. 1925. The effect of a small audience upon eye-hand coordination. *The Journal of Abnormal and Social Psychology* 20, 2 (1925), 142.

[57] Norman Triplett. 1898. The dynamogenic factors in pacemaking and competition. *The American journal of psychology* 9, 4 (1898), 507–533.

[58] Sarah Woods, Kerstin Dautenhahn, and Christina Kaouri. 2005. Is someone watching me?-consideration of social facilitation effects in human-robot interaction experiments. In *2005 international Symposium on Computational intelligence in Robotics and Automation*. IEEE, 53–60.

[59] Robert B Zajonc. 1965. Social facilitation. *Science* 149, 3681 (1965), 269–274.

[60] Catherine Amine Zanbaka, Amy Catherine Ulinski, Paula Goolkasian, and Larry F Hodges. 2007. Social responses to virtual humans: implications for future interface design. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1561–1570.

[61] Jakub Złotowski, Diane Proudfoot, Kumar Yogeeswaran, and Christoph Bartneck. 2015. Anthropomorphism: opportunities and challenges in human–robot interaction. *International journal of social robotics* 7, 3 (2015), 347–360.